**INFERENCES**

**>** prior\_summary**(**fit**)**

prior class coef group resp dpar nlpar lb ub source

normal**(**0, 10**)** b user

normal**(**0, 10**)** b alcohol **(**vectorized**)**

normal**(**0, 10**)** b price\_log10 **(**vectorized**)**

normal**(**0, 10**)** b Rich **(**vectorized**)**

normal**(**0, 10**)** b tannin **(**vectorized**)**

normal**(**0, 10**)** b tfidf\_tsne\_1\_norm **(**vectorized**)**

normal**(**0, 10**)** b tfidf\_tsne\_1\_norm**:**tfidf\_tsne\_2\_norm **(**vectorized**)**

normal**(**0, 10**)** b tfidf\_tsne\_2\_norm **(**vectorized**)**

student\_t**(**3, 0, 2.5**)** Intercept default

student\_t**(**3, 0, 2.5**)** sd 0 default

student\_t**(**3, 0, 2.5**)** sd variety 0 **(**vectorized**)**

student\_t**(**3, 0, 2.5**)** sd Intercept variety 0 **(**vectorized**)**

INFERENCES

* The model has assumed that the random effect intercept follows a Student's T distribution.

**>** summary**(**fit**)**

Family**:** bernoulli

Links**:** mu **=** logit

Formula**:** superior\_rating **~** 1 **+** price\_log10 **+** Rich **+** tannin **+** alcohol **+**

tfidf\_tsne\_1\_norm **\*** tfidf\_tsne\_2\_norm **+** **(**1 **|** variety**)**

Data**:** wine\_reviews **(**Number of observations**:** 2500**)**

Draws**:** 4 chains, each with iter **=** 2000; warmup **=** 1000; thin **=** 1;

total post**-**warmup draws **=** 4000

Multilevel Hyperparameters**:**

**~**variety **(**Number of levels**:** 35**)**

Estimate Est.Error l**-**95% CI u-95% CI Rhat Bulk\_ESS Tail\_ESS

sd**(**Intercept**)** 0.73 0.16 0.48 1.09 1.00 1072 1630

Regression Coefficients**:**

Estimate Est.Error l**-**95% CI u-95% CI Rhat Bulk\_ESS Tail\_ESS

Intercept **-**12.76 0.78 **-**14.30 **-**11.23 1.00 3327 3134

price\_log10 5.63 0.28 5.08 6.19 1.00 3924 2924

Rich 0.83 0.12 0.59 1.06 1.00 4334 2747

tannin 2.24 0.69 0.92 3.58 1.00 4741 3080

alcohol 2.72 0.57 1.59 3.83 1.00 4981 3197

tfidf\_tsne\_1\_norm 0.57 0.82 **-**1.08 2.16 1.00 2202 2447

tfidf\_tsne\_2\_norm 3.01 0.77 1.50 4.52 1.00 2160 2616

tfidf\_tsne\_1\_norm**:**tfidf\_tsne\_2\_norm **-**1.26 1.46 **-**4.06 1.58 1.00 2134 2481

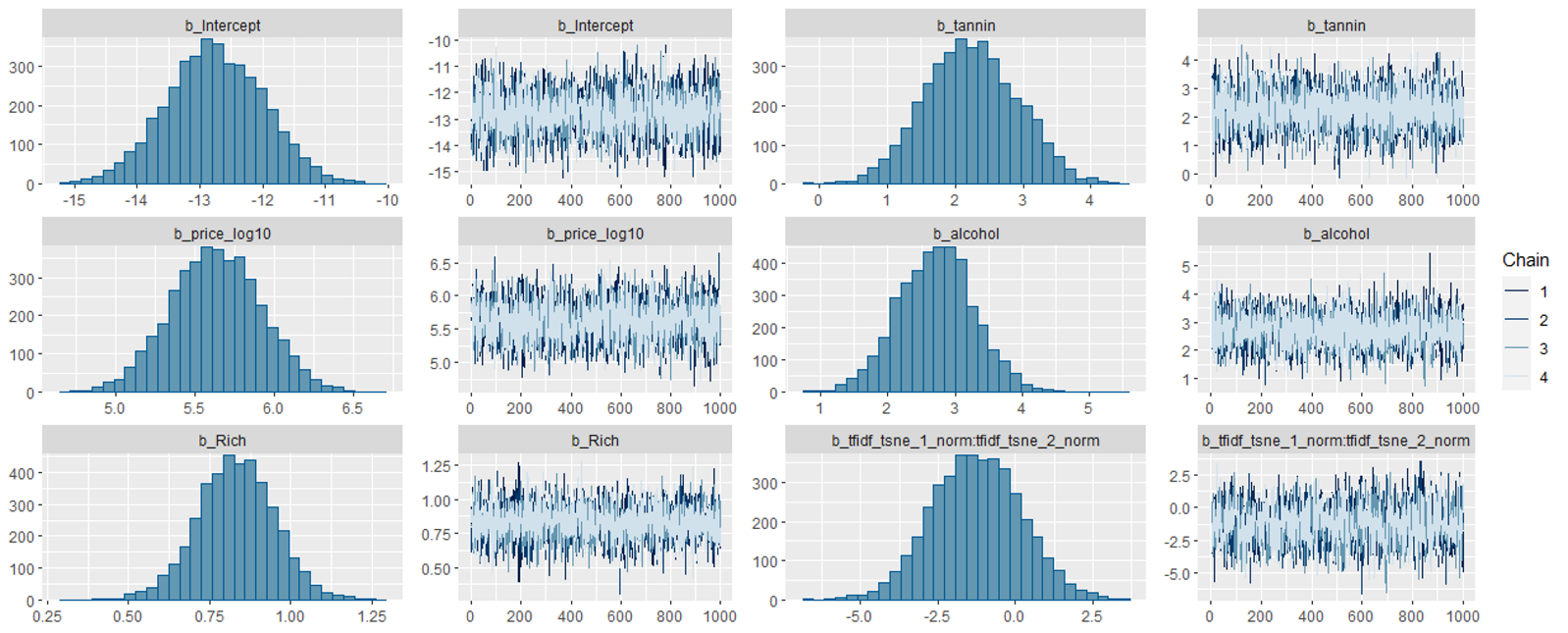
Draws were sampled using sampling**(**NUTS**)**. For each parameter, Bulk\_ESS

and Tail\_ESS are effective sample size measures, and Rhat is the potential

scale reduction factor on split chains **(**at convergence, Rhat **=** 1**)**.

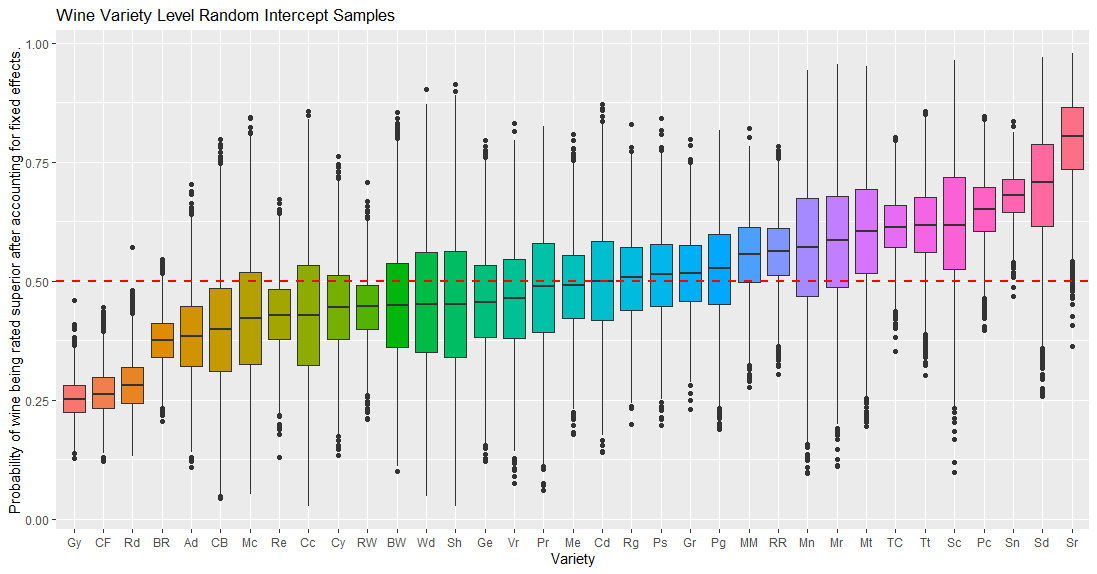
INFERENCES

* The posterior estimate of the standard deviation for the random intercepts (0.73 with an uncertainty of 0.16) suggests some, but not necessarily a vast amount of, variability in the log-odds of a wine being rated as "superior" across different varieties. This indicates that there might be some varieties that tend to receive consistently higher or lower ratings compared to the average. However, it's important to remember that this variability is on the log-odds scale. When translated back to the probability scale (by exponentiating), a standard deviation of 0.73 can translate to a noticeable difference in the probability of a "superior" rating between varieties.
* Rhat being 1 throughout, suggests that MCMC chains have converged well.
* Since ESS values are all much > 100, autocorrelation may be considered as being low enough to be acceptable.
* Of all the fixed effects in the model, the variable price\_log10 has the highest average effect on the log-odds of a wine being rated as "superior". Thus, it is likely a good estimator of whether a wine was rated highly or not.
* A one-unit increase in price (log transformed) is associated with an increase in the log-odds of being "superior" by an estimated 5.63 units, with a 95% credible interval of 5.08 to 6.19. This suggests a positive association between price (on a log scale) and being rated as "superior".
* The interaction term (tfidf\_tsne\_1\_norm: tfidf\_tsne\_2\_norm) has a negative coefficient and a wide credible interval, hinting at a complex relationship between the two transformed text features and the probability of a "superior" rating. This supports how the clusters observed upon plotting these 2 variables were really mixed.
* The model predicts a very low log-odds (-12.76) of a wine being rated as "superior" when all other factors are zero. This hints at a baseline tendency for wines to not get rated as being "superior". That said, the presence of the random intercept for "variety" with a considerable standard deviation (as indicated by the estimate and uncertainty) highlights the importance of variety-specific effects and suggests that some wine varieties might still have a higher chance of being rated "superior" compared to others.
* Both derived features, alcohol, and tannin, appear to be good predictors of superior ratings. The model estimates that a one-unit increase in alcohol content is associated with an increase of approximately 2.72 units in the log-odds of being rated "superior" (with some uncertainty). Similarly, a one-unit increase in tannin is associated with an increase of about 2.24 in the log-odds. Exponentiating these coefficients (e.g., ), translates to a potentially substantial increase in the probability of a wine being rated "superior" with higher alcohol content. The fact that these derived features have a larger effect size compared to the single indicator "Rich" reinforces the value of having engineered features to represent key wine judgement criteria while incorporating more than one wine characteristic.
* Plotting the distribution of the random intercepts across varieties might help get a better sense of the variability.



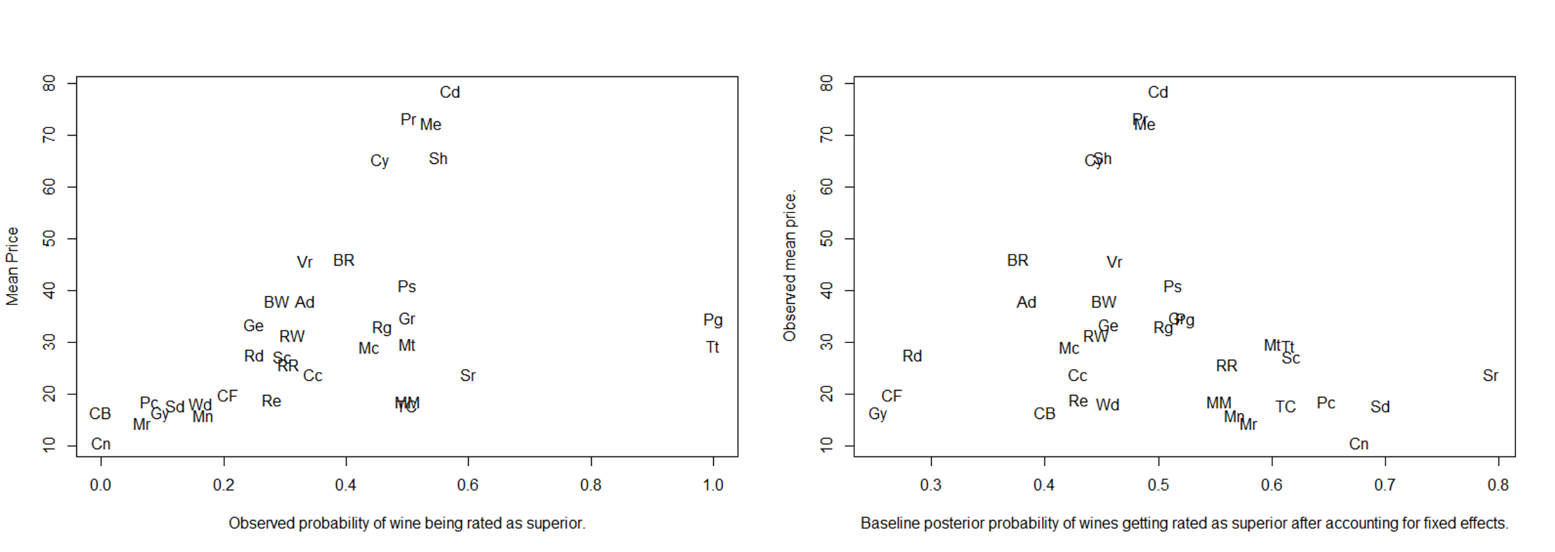
INFERENCES

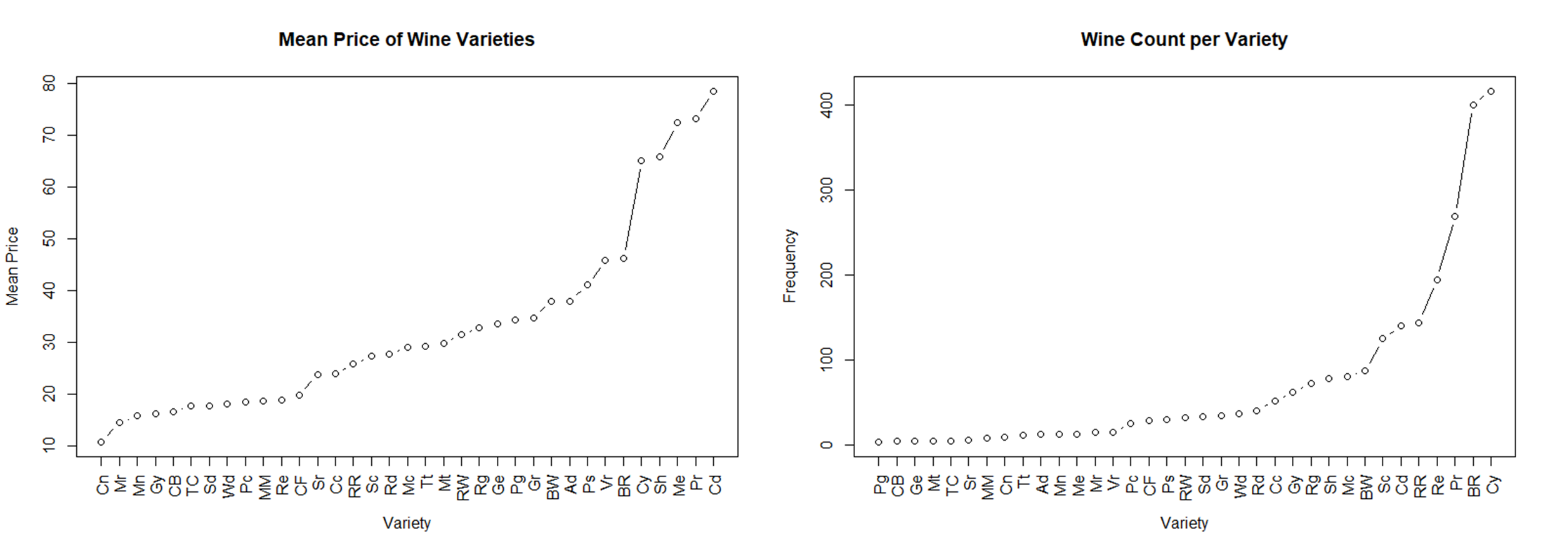
* All trace plots appear akin to random noise. This suggests good mixing and random MCMC sampling with low bias covering most of the probable regions, evenly. All chains arrived at the same random pattern indicating reliable convergence to stationarity suggesting that the target distribution was reached. Thus, we can safely conclude that here, the parameter space has been sufficiently explored and that posterior distributions represent reliable estimates.
* The posterior distribution visualized results from the summary table obtained previously. For instance, based on the credible interval, there is a 95% probability that the true value of the coefficient for price on a log scale lies within the range of 5.08 and 6.19, thereby suggesting a positive association between price and the log-odds of a "superior" rating. Converting from log scale to probability scale, here, a one unit increase in price would lead to an increase in the probability of wine being rated "superior" by a factor of anywhere between and times. At first glance, these numbers seem super large. But upon deeper analysis, this is plausible because this increase of 161 – 488 is relative increase w.r.t some baseline and not absolute change. So, if the baseline probability of a wine being rated "superior" was low, say 0.01, even a 180-fold increase would only be a small absolute increase in probability . Because the population parameter intercept was very negative (around -12) in the summary, it might be that baseline probability really is low.
* Generally, if a distribution is centered away from 0 and the credible interval does not have 0, it suggests evidence for a non-zero effect of that variable. That is, the further the distribution is from 0 and the narrower the credible interval, the stronger the evidence. [1] [2] Bearing this in mind, it can be sees here that barring the joint effect of tfidf\_tsne\_1\_norm and tfidf\_tsne\_2\_norm, all others seem to be having significant evidence against zero effect. That is, the chance that relationships (positive/negative) observed here between tannin content in wine, alcohol content, richness of the wine, its price, and the log likelihood of it having an associated superior rating, is less likely to be due to random chance. Of the predictors tried, the best ones (effect least due to chance = more likely to be a real effect) seem to be the price of the wine, its alcohol content and quantity of tannins in it.



INFERENCES

* Each box in the plot represents the distribution of posterior draws for a specific wine variety.
* Had the intercepts been directly plotted on the y axis, this would have indicated the per-variety baseline log-odds of a wine being rated "superior" after accounting for the fixed effects in the model due to the predictor variables. Since that might be harder to readily infer from, the log-odds were converted to probabilities prior to plotting using the formula . Thus now, the y-axis indicates the baseline probability that a wine of a particular variety gets rated as "superior" after accounting for the fixed effects in the model due to the predictor variables.
* The median line of each boxplot indicates the central tendency of the effect of that variety on the intercept.
* The box represents the interquartile range (IQR) of the distribution which contains the middle 50% of the posterior draws. The bottom edge of the box marks the first quartile (Q1) and the top edge marks the third quartile (Q3).
* The whiskers extend from the box towards the tails of the distribution such that they extend to the most extreme data points that lie within from the quartiles. Values outside this, are considered outliers.
* The 3 wine varieties at the leftmost end of the figure (“Gamay”, “Cabernet Franc”, “Red Blend”) are the ones with the lowest probability of having wines be rated as superior with minimal variance (more certainty).
* Varieties before the “Pg” variety (“Petit Manseng”), all have median intercept value below/very close to 0.5, meaning that consequently suggests that the odds of these varieties of wines being rated high enough to be ranked as superior is low.
* The wine variety with the highest odds of being rated as superior is “Sylvaner” on the extreme right of the figure. It’s five number summary values as well as that of neighboring varieties (“Sparkling Blend”, “Cabernet Sauvignon”, “Pinot Blanc”, …, “Tannat-Cabernet”) are all higher than 50% probability and thus have discernably higher odds of being ranked as superior to the varieties associated with the boxplots at the extreme left end. Thus, one may conclude that if one was to try wine of the “Sylvaner”, (“Sparkling Blend”, “Cabernet Sauvignon”, or “Pinot Blanc” variety, it’s very likely that this was a highly rated wine. Overall, the results show that some wine varieties do indeed get ranked high more often than some others with few having very low odds of being ranked as superior.
* That said, wine ranking is based on extremely subjective views. This likely explains the relatively wide spread of distributions in most cases. It is also possible that opinions on some variety, especially the ones in the middle of the image with some of the largest spreads, like “White Blend” and “Syrah” are very divided with both ardent fans and tough critics.





INFERENCES

* Comparing the proportion of wines that were rated superior as per observed data and the baseline posterior probability of wines getting rated as superior based on the random effects intercept, there seems to be an interesting effect. Many varieties that had a higher proportion of superior rated wines in the observed data was assigned a lower baseline probability after fitting the hierarchical logistic regression model. Some varieties like “Cabernet Sauvignon” (Sn) that had 9 associated data points in the data set with none of them being ranked superior, have been associated with high baseline log-odds of getting ranked as superior. This is likely because, the comparison here is apples to oranges. In the figure, the right plot shows observed proportion of "superior" wines for each variety in the data, but the left plot shows the baseline probability (converted from log-odds) of a wine being rated "superior" for each variety based on the model's estimates such that the x-axis reflects the grape variety's influence on the intercept (baseline log-odds) after accounting for fixed effects in the model.